

Retail Sales & Customer Behaviour Analysis

- **Project Background**

DataVista Pvt. Ltd. has been engaged by a mid-sized retail organization to conduct a comprehensive analysis of its customer purchasing behavior, product performance, and overall revenue trends. The objective of this engagement is to leverage data-driven insights to support strategic business decisions. Through detailed analysis, the management aims to identify opportunities to improve sales performance, gain a deeper understanding of customer preferences and buying patterns, optimize inventory planning and product assortment, and ultimately drive sustainable revenue growth across the business.

- **Dataset Structure**

1. customers.csv

Column	Description
CustomerID	Unique customer ID
CustomerName	Customer name
City	City of customer
Age	Age
Gender	Male/Female
SignupDate	Date customer joined

2. products.csv

Column	Description
ProductID	Unique product ID
ProductName	Product name
Category	Electronics, Fashion, Grocery, etc.
UnitPrice	Price per unit

3. orders.csv

Column	Description
OrderID	Unique order ID
CustomerID	Linked customer
OrderDate	Order date
PaymentMode	UPI / Card / Cash / NetBanking

4. order_items.csv

Column	Description
OrderItemID	Unique item ID
OrderID	Linked order
ProductID	Purchased product
Quantity	Units purchased
DiscountPercent	Discount applied

- **Intern Tasks**

Task 1: Data Understanding & Cleaning

The first phase of the project focuses on developing a clear understanding of the available datasets and ensuring data quality. Interns are required to load all provided CSV files and examine them for missing values, duplicate records, and inconsistencies. Appropriate data cleaning techniques should be applied to handle null values and remove or justify duplicates. Data types for all columns must be validated and corrected where necessary to ensure accurate analysis. As part of data preparation, interns must create a calculated field for Revenue, defined as:

$$\text{Revenue} = \text{Quantity} \times \text{UnitPrice} \times (1 - \text{DiscountPercent} / 100)$$

This calculated metric will be used throughout the analysis.

Task 2: Exploratory Data Analysis

In this phase, interns will perform exploratory data analysis to uncover patterns, trends, and relationships within the data. The analysis should answer key business questions, including the total revenue generated over the given period, monthly sales trends, and identification of the top five selling products. Interns should determine the best-performing product category, calculate the average order value, identify the most preferred payment mode, and analyze revenue distribution across different cities. The focus should be on deriving meaningful insights that can support management decision-making.

Task 3: Customer Analysis

The customer analysis task aims to understand customer demographics and purchasing behavior. Interns should differentiate between new and returning customers, analyze customer distribution across different age groups, and evaluate city-wise customer counts. Additionally, gender-based purchasing trends should be examined to identify differences in buying patterns and preferences among customer segments.

Task 4: Product Performance Analysis

This task focuses on evaluating product-level performance. Interns must analyze the impact of high versus low discounts on sales and revenue, identify products that generate the highest revenue, and detect low-selling products that may pose inventory risks. The objective is to assess product profitability and support inventory optimization decisions.

Task 5: Data Visualization

Interns are required to present their findings using appropriate visualizations created with tools such as Matplotlib, Seaborn, Power BI, or Excel. The visualizations must include revenue by category (bar chart), monthly revenue trends (line chart), city-wise sales distribution (bar or pie chart), payment mode usage (pie chart), and the top ten products by revenue (bar chart). All charts should be clear, well-labeled, and suitable for business presentations.

Task 6: Business Insights & Recommendations

In the final phase, interns must translate analytical findings into actionable business insights. Based on the analysis, they should recommend which cities should be prioritized for marketing efforts, identify the most profitable product categories, evaluate whether discount strategies should be increased or reduced, and suggest products that may need to be discontinued due to poor performance. Recommendations should be data-backed, clearly explained, and aligned with business objectives.

• Intern Submission Guidelines

At the completion of the project, interns are required to submit a complete project package consisting of a cleaned and well-structured dataset, the working analysis file (Python Jupyter Notebook, Excel workbook, or Power BI report), and a detailed insight report in PDF or DOC format. The report must clearly explain the data cleaning steps, analysis approach, key findings, visualizations, and business recommendations derived from the dataset. Interns must also include screenshots of all major charts and dashboards created during the analysis to demonstrate their understanding of data visualization techniques. Additionally, all project files must be uploaded to GitHub or Google Drive, and the shareable link should be included in the final submission. For confidentiality and industry compliance reasons, actual product names and sensitive business identifiers must be masked or replaced with generic labels throughout the report, dataset, and visualizations.

• Report Writing Instructions

The project report should be structured in a professional business format and must include the following sections:

- 1. Project Overview**
Briefly explain the business problem, project objective, and the role of data analytics in solving it.
- 2. Dataset Description**
Describe the datasets used, table relationships, and key columns.
Note: Product names were anonymized in dataset for confidentiality.
- 3. Data Cleaning & Preparation**
Explain how missing values, duplicates, incorrect data types, and inconsistencies were handled.
- 4. Exploratory Data Analysis (EDA)**
Present trends, patterns, and distributions using charts and summary statistics.

5. **Key Insights & Findings**

Highlight important observations related to customer behavior, sales performance, and product trends.

6. **Business Recommendations**

Provide actionable, data-driven suggestions that management can use for decision-making.

7. **Conclusion**

Summarize outcomes and mention possible future improvements or advanced analysis.